The Virtual Magnetospheric Observatory at UCLA

Raymond J. Walker, Todd A. King, Steven P. Joy, Lee F. Bargatze, Peter Chi, James M. Weygand, Robert L. McPherron and Christopher T. Russell
University of California, Los Angeles

Abstract

The successful management of space science observations requires a combination of expertise in both science and information technology. The Virtual Magnetospheric Observatory (VMO) creates robust links to the world's relevant data bases thus providing one-stop shopping for the magnetospheric researcher seeking data. The VMO is being implemented by using existing and widely adopted technologies and strives to provide well organized views of diverse science holdings. Since data are very dynamic especially during the early phases of a mission, the VMO portal design allows frequent and asynchronous updating. The VMO portal provides access to value-added services (e.g. to reformat, manipulate, analyze and display data) developed both locally and remotely. The registries for both data and services are designed to make it easy for suppliers to make their resources available and update information regarding their resources. The basis for resource descriptions is the SPASE data model. We have created tools to enable data repositories to communicate with the VMO even if they use other data models. Finally we work with data suppliers to create archival quality data products.

The Problem

Our Approach

With our colleagues at VMO-G :

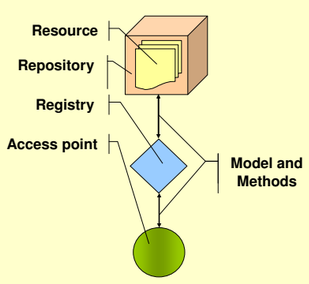
- Provide a single request point (portal) for all data repositories needed for magnetospheric research.
- Provide tools for preparing data archives.
- Provide services or access to existing services to reformat, manipulate, analyze and display data.
- Identify and register magnetospheric resources.
- Expand data model to support data processing (services) and data quality.

The Self-Organizing Data System

- The Data System is a collection of repositories.
  - Archives are organized according to a data model.
  - Repositories typically are distributed.
- The first step in developing the self-organized data system is adopting a common data model and dictionary.
  - Metadata are the glue that binds distributed systems.
  - Metadata adheres to an organization defined in a data model.
  - The SPASE (Space Physics Archive Search and Extract) consortium is developing a candidate data model and dictionary for space physics.
  - The data model is being defined by scientists.
- Repositories document their resources according to the data model by using the shared data dictionary.
  - Resources include data, documentation, software, services, etc.
  - Each resource has a metadata envelope that allows it to be part of a bigger system.
  - Repositories grow resource by resource at their own rate.
- The Data System self-organizes based on the metadata.
  - Repositories are concerned only about resources under their purview.
  - The Data System is concerned about how these resources are semantically related.
- All metadata are stored in registries.
  - Typically registries are co-located with the repository, but they can reside at a remote site.
- The Data System queries all registries to generate a response.
  - A query to the system will always return the most complete answer, even with dynamic repositories.

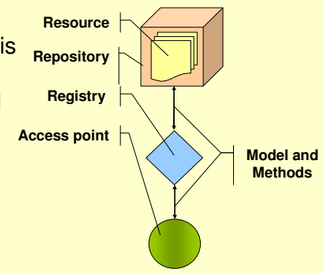
Components of VMO

- **Resource**: An object (document, data, etc.) or service available for use.
- **Repository**: A facility for storing and maintaining digital information in accessible form.
- **Registry**: A collection point for metadata about resources.
- **Access Point**: An interface to the registries and resources.

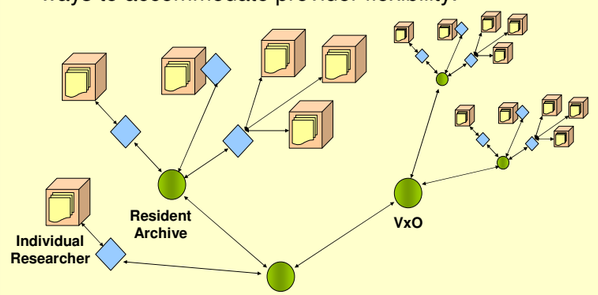


The Fundamental Building Blocks

- **Data Model**: Describes in an abstract way how data is represented. This includes semantics (meaning of terms) and ontology (relationships between terms).
- **Access Methods**: Mechanisms to search for, use, and distribute resources.



- The building blocks can be combined in many ways to accommodate provider flexibility.

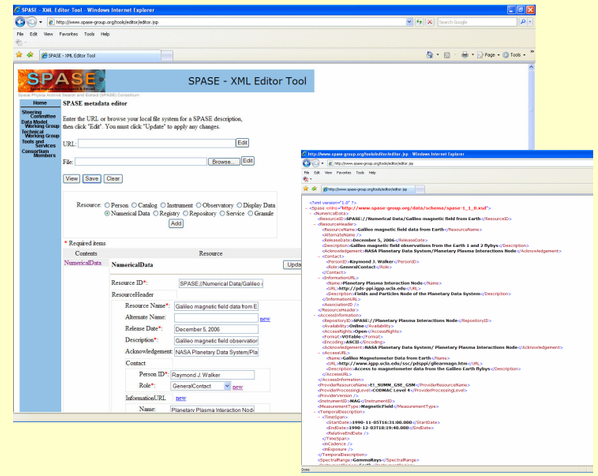


Steps in Building the VMO

1. Define the data model: (SPASE)
   - Select or write the data dictionary.
   - Create the ontology of terms (relationships among terms).
2. Choose a standard representation for expressing the relationships (XML).
3. **Describe resources**
   - **Use tools to help large scale tasks (editor and rule sets).**
4. Populate the registry
   - Obtain data descriptions for each repository.
5. Deploy services to search and extract data.
   - Interfaces to the registry for searching.
   - Interfaces to the repository for extraction.
6. Value added services.
   - Joining and sub-setting.
   - Transformation (reformatting, coordinate transformations, etc.).
   - Display.

Tools to Describe Resources

- SPASE editor aids repository staff in creating SPASE compliant XML.



- Rule set processing automates metadata generation
  - Rule set processing uses a platform independent language to collect and format metadata.
  - A rule set is a collection of rules – a rule is a statement of a needed action (e.g. assign a value to a variable, include another rule set, run an external application, write output, etc.)
  - The rule set language allows "plug-in" routines that extract information from the data, process it and write the results out.
  - Rule sets can be written by data providers or by VMO scientists.
  - The providers can then use rule sets to prepare their resource descriptions with minimal effort.